Tooth Growth Exploratory Data Analysis - Statistical Inference Assignment Part 2

Nathaniel Landi

Sunday, September 21, 2014

The Toothgrowth data set available from R's dataset bundle is used to perform a basic exploratory analysis with the various statistical concepts learned in the class.

1. Loading the dataset

The dataset library contains the Toothgrowth dataset and is loaded and attached to the working directory for ease of use with calling the variables by name.

```
library(datasets)
library(ggplot2)
library(psych)
```

```
##
## Attaching package: 'psych'
##
## The following object is masked from 'package:ggplot2':
##
##     %+%
```

```
library(gridExtra)
```

```
## Loading required package: grid
```

```
library(xtable)
data(ToothGrowth)
attach(ToothGrowth)
```

The various columns in the data set are analysed and the data contained shows the effect of Vitamin C on Tooth growth in guinea pigs. The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods OJ and VC (orange juice or ascorbic acid). The data has 60 observations on the three variables.

```
# Columns in the data set
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```r
# Statistical Summary of the data set
summary(ToothGrowth)
```

```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

The distribution of the observations are found to be equal.

```r
# Frequency of the data eliminating missing Values
table(supp,dose)
```

```
##     dose
## supp 0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

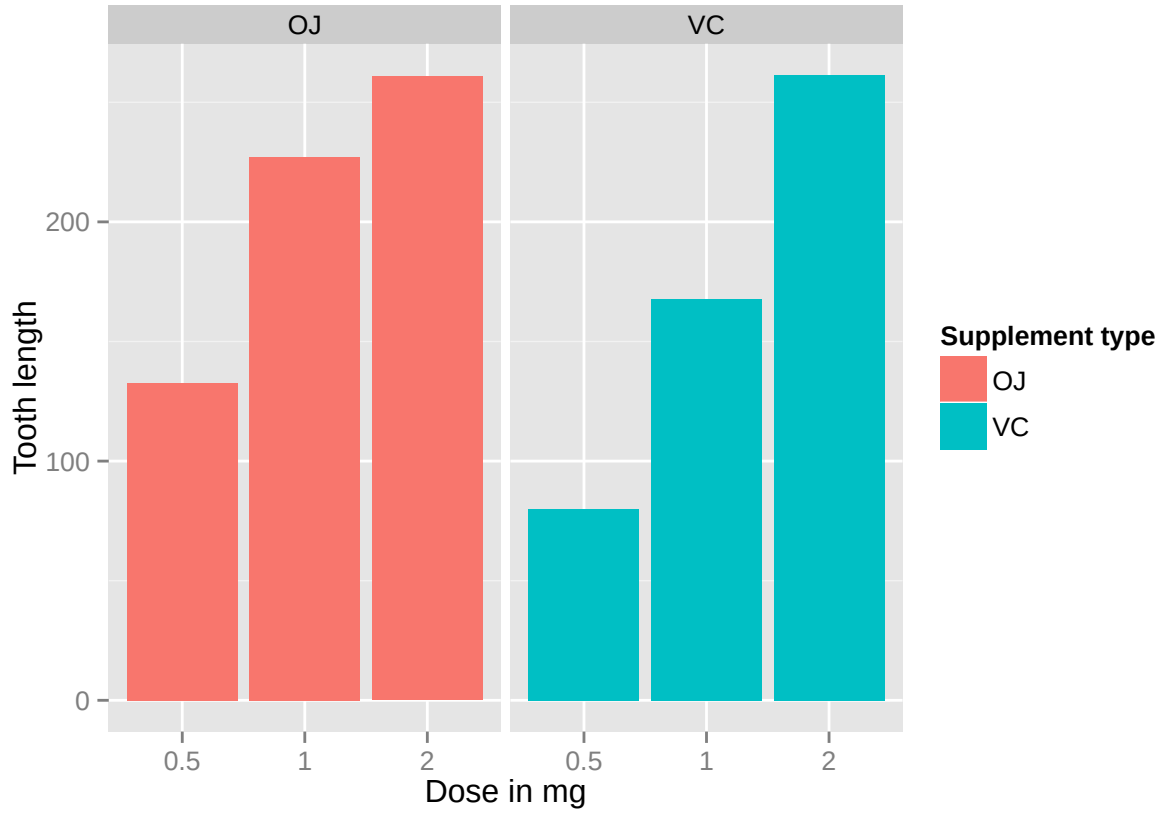
The overall dataset statistic is also calculated.

```r
# Summary statistics by grouping as variables
dose <- as.factor(dose)
describe(len)
```

```
##   vars  n  mean   sd median trimmed  mad min  max range  skew kurtosis
## 1    1 60 18.81 7.65  19.25   18.95 9.04 4.2 33.9  29.7 -0.14    -1.04
##     se
## 1 0.99
```

The data is plotted as a bar plot to identify any trends between the variables.

```r
# Correlation between Supplement on Tooth Growth
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_bar(stat="identity",) +
  facet_grid(. ~ supp) +
  xlab("Dose in mg") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```

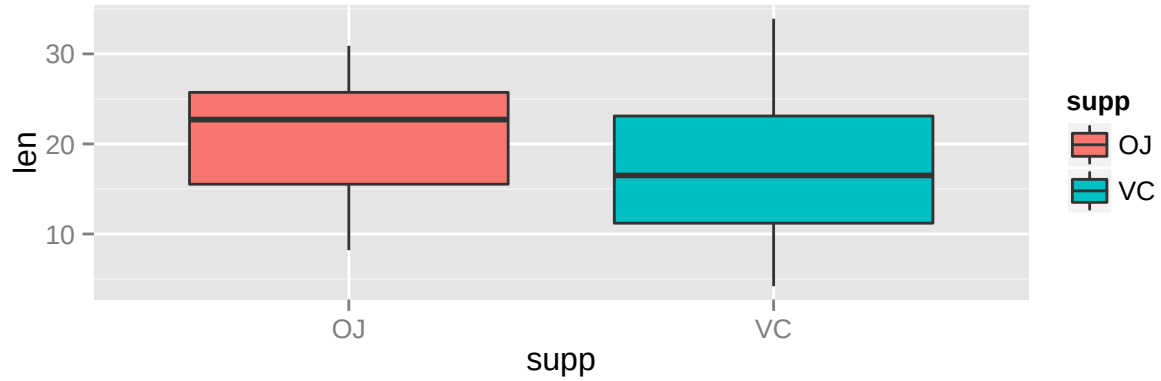


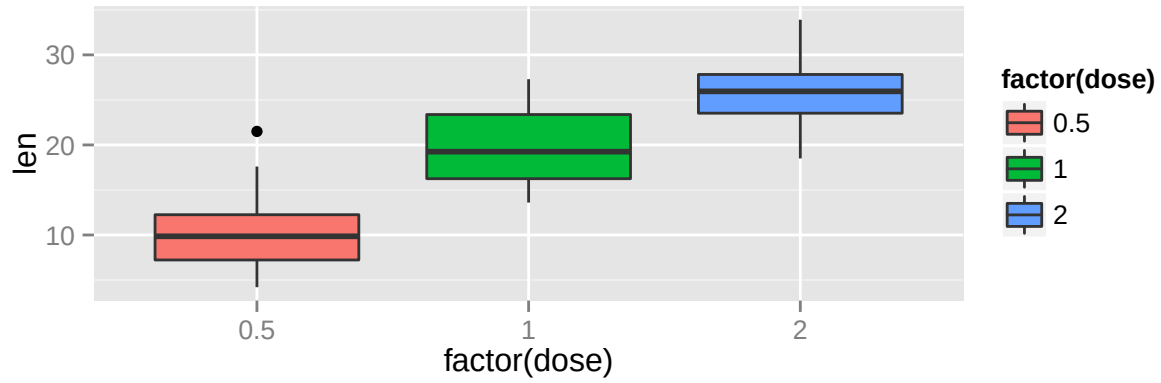
It can be seen from the plot that as the dosage of vitamin C is increased there is an increase in tooth length. This direct correlation is more pronounced when ascorbic acid is used as a vector for delivery for all three doses (0.5, 1, 2 mg) but when orange juice is used as a vector for delivery, the effect is prominent in the lower dose levels but is not very significant in the increase from 1 to 2 mg dose levels.

To further identify the effect of dosage on tooth growth, two box plots were plotted with using just the difference between the supplement and just the dose levels.

```
# Effect of Dosage on Tooth Growth
suppplot<-ggplot(aes(x = supp, y = len), data = ToothGrowth) + geom_boxplot(aes(fill = supp))
doseplot<-ggplot(aes(x = factor(dose), y =len), data = ToothGrowth) +
  geom_boxplot(aes(fill = factor(dose)))

grid.arrange(suppplot,doseplot)
```





It can be seen that Orange juice has higher potential as vitamin C delivery vehicle than ascorbic acid with higher average tooth growth and also the increase in dosage does increase the tooth growth but the growth ratio is from 0.5 to 1 vs 1 to 2 is significantly different. The mean and standard deviation are calculated for each of the doses before performing a sample t-test to justify or nullify the hypothesis that supplements cause tooth growth.

```
# Mean length for the two supplements
round(with(ToothGrowth, sapply(split(len, supp), mean)), 3)
```

```
##    OJ    VC
## 20.66 16.96
```

```
# Mean length for the two supplements for various doses
aggregate(len, list(supp, dose), mean)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 13.23
## 2      VC     0.5  7.98
## 3      OJ       1 22.70
## 4      VC       1 16.77
## 5      OJ       2 26.06
## 6      VC       2 26.14
```

```r
# Standard Deviation for the two supplements for various doses
aggregate(len, list(supp, dose), sd)
```

```
##   Group.1 Group.2     x
## 1      OJ     0.5 4.460
## 2      VC     0.5 2.747
## 3      OJ       1 3.911
## 4      VC       1 2.515
## 5      OJ       2 2.655
## 6      VC       2 4.798
```

Performing a regression analysis will help justify the effect of the dose and also address the issue whether the supplement type has an effect by measuring the variance if any caused by the choice of supplement.

```r
# Regression to see the effect of supplements on growth
fit <- lm(len ~ dose + supp, data=ToothGrowth)
summary(fit)
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.273      1.282    7.23  1.3e-09 ***
## dose           9.764      0.877   11.14  6.3e-16 ***
## suppVC        -3.700      1.094   -3.38   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.24 on 57 degrees of freedom
## Multiple R-squared:  0.704,  Adjusted R-squared:  0.693
## F-statistic: 67.7 on 2 and 57 DF,  p-value: 8.72e-16
```

The regression results show that there is a 70% variance in the data and the average length of the teeth without any supplements is 9.725 units in length. Given that the coeffient of dose is 9.7636 units, a 1 mg increase in dosage will increase the length by 9.7636 units. Also interesting to note is the fact that the VC supplement has a negative coefficient implying the orange juice has a positive effect on teeth growth. The low p-value also suggests to reject the null hypothesis suggesting that each variable can address a significant portion of variability in the growth of the tooth.

Hypothesis testing

```r
# T-Test to find correlation between Orange Juice vs Vitamin Supplement
```

```
dose_0.5 <- ToothGrowth[ToothGrowth$dose == 0.5,]
dose_1 <- ToothGrowth[ToothGrowth$dose == 1,]
dose_2 <- ToothGrowth[ToothGrowth$dose == 2,]

Testdose_0.5 <- t.test(dose_0.5$len~dose_0.5$supp , var.equal = T)
Testdose_1 <- t.test(dose_1$len~dose_1$supp , var.equal = T)
Testdose_2 <- t.test(dose_2$len~dose_2$supp , var.equal = T)

conf_0.5 <- Testdose_0.5$conf[1:2]
conf_1 <- Testdose_1$conf[1:2]
conf_2 <- Testdose_2$conf[1:2]

pvalue_0.5 <- Testdose_0.5$p.value
pvalue_1 <- Testdose_1$p.value
pvalue_2 <- Testdose_2$p.value
```

The p values and confidence interval for doses of 0.5 mg:

```
pvalue_0.5
```

```
## [1] 0.005304
```

```
conf_0.5
```

```
## [1] 1.77 8.73
```

The p values and confidence interval for doses of 1.0 mg:

```
pvalue_1
```

```
## [1] 0.0007807
```

```
conf_1
```

```
## [1] 2.841 9.019
```

The p values and confidence interval for doses of 2.0 mg:

```
pvalue_2
```

```
## [1] 0.9637
```

```
conf_2
```

```
## [1] -3.723  3.563
```

We can see that there is a significant difference between the means of the supplements when the dosage levels were 0.5mg and 1 mg for Orange juice vs Ascorbic acid but the effects are interchanged for the 2 mg dosage level. The p-values also suggest that at 0.5 mg and 1 mg dosage level, orange juice is statistically significant carrier than ascorbic acid while at the 2 mg level, the evidence is statistically insignificant.

Conclusion

Basic exploratory analysis of the toothgrowth data reveals that supplements help in tooth growth in guinea pigs and with a high level of confidence, it can be stated that orange juice acts as an effective vector for tooth growth in doses of 0.5 mg and 1 mg compared to ascorbic acid.